

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 4, 2003, 12:21:43 ; Search time 331 Seconds

(without alignments)
14019.136 Million cell updates/sec

Title: US-10-082-894-1

Perfect score: 1719

Sequence: 1 gtttaattaccgaatttga.....aaatctcgtataaaaatt 1719

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_19Jun03:*

- 1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
- 23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
- 24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
- 25: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	332.8	19.4	1386	21	Aspergillus oryzae
C	314.2	18.3	684707	24	Listeria innocua c
C	314.2	18.3	3011208	24	Listeria innocua D
C	312.6	18.2	2944528	24	Listeria monocytog
C	284.8	16.6	580073	18	Mycoplasma genital
C	271.2	15.8	1554	24	Staphylococcus epi
C	269.6	15.7	3373	22	S. epidermidis gen
C	266.2	15.5	1518	22	S. epidermidis ope

9	243	14.1	6048	18	AAV74439
10	231.2	13.4	65792	22	AAF28544
11	222.2	12.9	1620	19	AAF14157
C	219.2	12.8	11769	24	ABS53606
13	175.6	10.2	1041	24	ABK73405
14	161.2	9.4	881	24	ABK77929
15	125.6	7.3	1280	23	AAS85747
16	112.2	6.5	1888	21	AAC41181
17	103.4	6.0	832	24	ABQ68729
18	103.4	6.0	4230	24	ABQ70965
19	98.2	5.7	1674	18	AAAT66245
20	98.2	5.7	1960	21	AAAT5770
21	95.2	5.5	2094	21	AAA26936
22	95.2	5.5	2094	22	AA08087
23	95.2	5.5	2094	22	AAF91618
24	92.2	5.4	294	25	AB255893
25	91.6	5.3	1963	21	AAAT1752
26	89.8	5.2	1671	18	AAAT6243
27	88.2	5.1	1668	18	AAAT6244
28	86.2	5.0	1683	18	AAAT6246
29	84	4.9	1728	21	AAAT6246
30	84	4.9	1728	21	AAAT6246
C	81.4	4.7	484	19	AAV69419
C	81.4	4.7	484	21	AAZ89984
C	81.4	4.7	484	24	ABS69845
C	81.4	4.7	484	24	ABS69845
35	80.8	4.7	1593	18	AAAT6242
36	69.4	4.0	400	18	AAV75093
37	68.6	4.0	254	24	ABX65706
38	61.6	3.6	1641	24	ABQ90454
39	54	3.1	294	24	ABN18433
40	53.4	3.1	732	24	ABQ45492
C	53.4	3.1	732	24	ABQ45493
C	51.8	3.0	264	24	ABK73603
43	46.8	2.7	183	24	ABN77799
44	45.2	2.6	736	21	AAAT6246
45	45.2	2.6	837	21	AAAT6246

ALIGNMENTS

RESULT 1

AAAT6246

ID AAF13546 standard; cDNA; 1386 BP.

AC AAF13546;

DT 13-MAR-2001 (first entry)

DE Aspergillus oryzae EST SEQ ID NO:6069.

XX Multiple gene expression; filamentous fungal cell; EST;

KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;

KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;

KW culture condition; environmental stress; spore morphogenesis;

KW metabolic pathway engineering; catabolic pathway engineering; ss.

OS Aspergillus oryzae.

XX WO200056762-A2.

XX 28-SEP-2000.

XX 22-MAR-2000; 2000WO-US07781.

XX 22-MAR-1999; 99US-0273623.

PA (NOVO) NOVO NORDISK BIOTECH INC.

PA (NOVO) NOVO NORDISK AS.

PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;

XX

WPI; 2000-594572/56.

Monitoring differential expression of genes in filamentous fungal cells uses fluorescence-labeled nucleic acids isolated from the cells and a substrate of expressed sequence tags -

Claim 88; Page 2502-2503; 3161pp; English.

The present invention describes a method for monitoring differential expression of genes in a first filamentous fungal (FF) cell relative to expression of the same genes in one or more second filamentous fungal cells. The method uses fluorescence-labeled nucleic acids isolated from the FF cells and a substrate of expressed sequence tags (EST). The ESTs are used in the methods for monitoring differential expression of genes in a first filamentous fungal (FF) cell relative to expression of the same genes in one or more second filamentous fungal cells. Monitoring the global expression of genes from FF cells allows the production potential of the microorganisms to be improved. New genes may be discovered, possible functions of unknown open reading frames can be identified and gene copy number variation and stability can be monitored. The expression of genes can be used to study how FF cells adapt to changes in culture conditions, environmental stress, spore morphogenesis, recombination, metabolic or catabolic pathway engineering. Using ESTs provides several advantages over genomic or random cDNA clones including elimination of redundancy as one spot on an array equals one gene or open reading frame, and organisation of the arrays based on function of the gene products to facilitate analysis of the results. AF07478 to AF11247 represents ESTs from *Fusarium venenatum*; AF11248 to AF11853 represents ESTs from *Aspergillus niger*; AF11854 to AF14878 represents ESTs from *Aspergillus oryzae*; and AF14879 to AAP15337 represents ESTs from *Trichoderma reesei*, which are all specifically claimed in the present invention.

Sequence 1386 BP; 317 A; 389 C; 383 G; 297 T; 0 other;

Query Match	19.4%;	Score 332.8;	DB 21;	Length 1386;
Best Local Similarity	57.7%;	Pred. No. 1.2e-76;		
Matches 694;	Conservative 0;	Mismatches 492;	Indels 16;	Gaps 5;
QY	181	CTTCATGTTGGATTGCGACGAAGGCTTAATGGGAAATTTCTGAAGTTGGACATTTGAATATA	240	
DB	197	CTTGCTGTGGATTACCGGAAGTCTCATGGGTAACAGTGAAGTCGCCATCTGAACATT	256	
QY	241	GGAGCTGGGAAGATTATTATCAAGATATTTGTTGCAATTAATTTGGCTGTTCAACGAAC	300	
DB	257	GGTGGGGAAGCTGTGCTGGCAAGACAGTGTTCGTATCGACCAAAACCCCTCAAGAAAGGC	316	
QY	301	GAGTTGTTTACAATCCTCAGATTGTTGCATCAGCTGAGCGTGCACAAAGAGGGAGTGGT	360	
DB	317	GAGTTCAACAAGGTACACACGTCGTTGTCATCTTCAAGCGTGCCAAAGGAGGCATGGC	376	
QY	361	CGATTGCATTTATTAGGACTGGTTAGCGATGGTGGTGTCCACTCTCATATTGATCATCTT	420	
DB	377	CGTCTTCATCTCTGGGCGCTGGTCTCCGACGGTGGTGTCCACTCCCAACATCACCCACCTT	436	
QY	421	TTTGCGCTGATACGTGCATTTAAACAATTACAAAGTGCCAAAGGTTTTTCATTCACGTTTTT	480	
DB	437	ATCGGCCCTGTTGAAGGTTGCGCAAGGAGATGGAGATTCCCAAGGTTTTTCATCCAGCTCTTC	496	
QY	481	GCTGATGGTCGATATCTTCGCCAACAAAGTGGAGCTGGTTATCTTGAACAACTCTCTTCAA	540	
DB	497	GCGATGGAGCTGACACCGNACCCAGAGCGCTACCAAGTACATGCAGAGCTTCTCGAT	556	
QY	541	TTTATTGCTTCGAAAAGTACGGAGAAATTTGGCTTACTATTACTGACGCTTATTATGCAATG	600	
DB	557	CAGACCAAGGAATCGGCATTTGGTGAATCGCCACTGTTGTTGGACGCTACTGGGCCATG	616	
QY	601	GATAGGCGAAAAAGATGGGAGCGCTATTAAAGATGCTTATGACGCAATGTTGGAGGATTT	660	
DB	617	GACCGCGACAGCGCTGGACCGCTGCGAATTTGCCATGAAGGCAATTTGTCGGGAGAG	676	
QY	661	GGACAAAAAGCCACCGCTTGATTAAGGCTTGTCGATGTTGTTAGAGAGCGGATATGCTCAATCT	720	

Db	677	GGCGAGGATCGTGGCATC-----CTGTTAAGACCATCAACGAGCGCTACGAGAAGGAT	730
Qy	721	GAGACTCAGCAATTTCTGAAACC---AATTGTTTTTTTCGGACGATGGGAGTAAGAAGAT	777
Db	731	GAGACCGACGAGTTCTGTGAAGCCCATCATCGTCGGTGGCGAGGAGACAGCGGTCAAGGAC	790
Qy	778	GAGCATACTCTTATTTCTCAANTATCGTGCTGATCGTATGGTCAAAATTTGCTGAATGT	837
Db	791	GATGACACCCCTTTCTTCTCAACTACCGCTCTGACCGGTTCGCGAAATACCCCAACTG	850
Qy	838	TTGGGCTCTCAACGGTTATAAAGATCTAATAGTTTCGGTTCCTCACCCCTCAAAAATATTCAG	897
Db	851	CTCGGT--GACTACGACGCGACGCCCTAAGCCCGACTTCCTCTTACCCGAGAACAATCCAC	907
Qy	898	ATTAGTGGGATGACCAATACATAAAGAGTTTCCATTTCCCATCGTTATTTCCCACTGTG	957
Db	908	ATTACCACTATGACCCGATATAAGCGACTACACATTCCTCTGTCTTTTCCCCCCTCAG	967
Qy	958	ACTCATACTAATGTGCTTGTGTAATGGCTTGTCTTCTCAAGGAGTTACTCAATTTCACTGT	1017
Db	968	CACATGGGTAACTGCTCGCTGAATGGCTCAGCAAGAGGATCTTCAGCAATCTCAGTT	1027
Qy	1018	GGCGAACTGAGAAAGTATCTCTATGTTCCTTCTTTAATGGTGGTCGAGAAGTTCAA	1077
Db	1028	GCCGAGACTTGAGAAGTACGCTACGTCACCTTTCTTCTTTAACGGTGGTATTGAGAAGCAG	1087
Qy	1078	TTCCAAAGATGAAGAGCGTTGTATGTGTTCCGTACCAAAAGAAGTGTGTACATGATTTTA	1137
Db	1088	TTTCGCTGGCGAGGTCCGCGACATGATCCCATCCCCAA--AGTTGGGACCTACGACCTG	1144
Qy	1138	AAACCGAAGATGAATGCTGTGGAGTTGCGGAAAAAATGGTCGAGCAAAATTGAGTCAGGC	1197
Db	1145	GACCCCAAGATGAGCGCTGAGGCGCTGGGTTCAGAAGATGGCCGACGCGATTTGCTGAGGC	1204
Qy	1198	AGGCATCCTTTGTTTATGTCAATTTTGGGCCCTCTGACATGTTTGGACATCTCGTAAA	1257
Db	1205	AAGTTCGAGTTCGTCATGAACAATCTTCGCTCTCTTCGACATGGGTGGCCACACTGGTAAG	1264
Qy	1258	TTTGAACCTGCCGTCAAAGCATGTCGAAGCTTACTGACGAGGCAATTTGGAAGATATTGAA	1317
Db	1265	TACGAAGCTGGCATTCAGGGTGTGTGCTTACCGAGAAGGCCATCGGTGTCTATGAA	1324
Qy	1318	G-CATGCCAAACTTATAATTAACGTTCTTATGTTTACTTCGGATCATGGAAATGCTCAGAA	1376
Db	1325	GCCTTCGAAGAACAAGGCTATGTGCTCTTCATCATCTGCGATCATGGAAACGCCAGGA	1384
Qy	1377	GA 1378	
Db	1385	AA 1386	

ABQ6/196/C
ID ABQ67196 standard: DNA: 684707 BP.

AC AB067196:

DT 29-AUG-2002 (first entry)

DE Listeria innocua contig DNA sequence #9.

Antibacterial; Listeria; food contamination; mutational analysis;
KW infection: ds.

AA
OS
Listeria innocua.

PN WO200228891-A2.

11-APR-2002.

PF 04-OCT-2001;

PR 04-OCT-2000; 2000FR-0012697.


```
XX PD 11-APR-2002.
XX PF 04-OCT-2001; 2001WO-FR03061.
XX PR 04-OCT-2000; 2000FR-0012697.
XX PA (INSP ) INST PASTEUR.
XX PA (CNRS ) CNRS CENT NAT RECH SCI.
XX PI Kunst F, Glaser P;
XX DR WPI; 2002-332479/37.
XX PT New genomic sequences from Listeria species, useful for detection,
XX PT treatment and prevention of infection, also related polypeptides,
XX PT antibodies and modulators
XX PS Claim 5; SEQ ID 2058; 180pp; French.
XX CC The present invention relates to nucleic acid sequences
XX CC (AB067188-AB071212) from Listeria sp. The sequences are useful as probes
XX CC and primers for identification and/or detection of Listeria (e.g. as
XX CC contaminants in foods, or mutational analysis) and for analysis of
XX CC gene expression. Proteins encoded by the nucleic acid sequences can be
XX CC used to screen for compounds that modulate gene expression, replication
XX CC and pathogenicity of Listeria (potential therapeutic agents), also for
XX CC treating infections by Listeria, and are useful as immunogens in
XX CC anti-Listeria vaccines.
XX CC Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic format
XX CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 3011208 BP; 941651 A; 568176 C; 559189 G; 942192 T; 0 other;

Query Match 18.3%; Score 314.2; DB 24; Length 3011208;
Best Local Similarity 53.3%; Pred. No. 1.9e-70;
Matches 767; Conservative 0; Mismatches 648; Indels 24; Gaps 4;

QY 14 AGTTTGAGATGACAAATATCAAAATGTTCAACAAAAGCTGCTGTGTAGTTATGATG 73
DB 2576453 ATATTTGGAGGAGTAAATAAATAGTAAATACCTGCTAGCAATATATCTCGATG 2576394
QY 74 GATGGGGCTTTCCGATGAACACACACGGGAATGCAATGCTAAAGCTAAAGCCCTATTA 133
DB 2576393 GTTTGGTAAACGTGCAGAAACAGTAGGTAATGCTGCTCAAGCAACAAACAAAT 2576334
QY 134 TCGACAAACTTTGTTCTGGAATTTGGCA---AAATTTGGAACACACGGTCTCATGTTG 190
DB 2576333 TCGACCGTTTACTGGGCTAATTTTCTCAGGGGAACTTAAAGCTGCTGGCCTTGATGTTG 2576274
QY 191 GATTGCGACAGGCTTAATGGGAAATTCCTGAAGTTGGACATTTGAATATAGGAGCTGGAA 250
DB 2576273 GTCCTCCAGAGCCCAATGGGTAACTCTGAAGTTGGCCATACAACTCGGAGCTGGAC 2576214
QY 251 GAGTTATTTATCAAGATATTTGTCGAATTAATTTGGCTGTTCAAGCAACGAGTTGTTA 310
DB 2576213 GTATTGCTTACCAAGCTTAACTCGTATTGATAAGCAATTCGAAGCGAATTCGAAG 2576154
QY 311 CAATTCCTCAGATTTGTCATCAGCTGACGCTGCAAGAGGAGGAGTGCTCAATTCGATT 370
DB 2576153 AGAACAAGGCCCTAAACAATGCTTTTCACTCATACAAAGAAACAAATTCGGACCTACATC 2576094
QY 371 TATTAGGACTGTTAGGAGTGTGTGCTCACTCTCATATTCATCATCTTTTGGTGTGA 430
DB 2576093 TTTTCGGTTACTATACAGCGCGGGTGTCATAGTCACATTAATCACTTCGTTCGACTTT 2576034
QY 431 TACGTGCAATTTAAACAATTTACAAGTGCCAAAGGTTTTTCATTCACATTTTTTGTGATGTC 490
DB 2576033 TAGAACTTCGGAAGATAAAGCGGTGAAAGAGTTTTATATCCATTCATTCCTTTGATGAC 2575974
QY 491 GAGATACTTCGCCAACAAAGTGAGCTGTTATCTTTGAACAACCTTCTCAATTTATGCTT 550
```

```
DB 2575973 GTGACGTGGCACCAACATCCTCACTAGATAATTTAGAAAACACTAGAAAAGCTATTAGTG 2575914
QY 551 CGGAAAGTACGAGGAATTTGGCTACTATTACTGGAGCTTATTTATCAATGGATAGGACA 610
DB 2575913 ACTTAAACTATGTCGAATTTGCTACTGTTCTTGGACGTTTCTACGGGATGGATCGTATA 2575854
QY 611 AAAGATGGAGCGTATTAAAGATGGCTTATGAGGCAATTTGTTGAGGATTTGGACAAAAAG 670
DB 2575853 AACGCTGGGACGTTGTGAAAAGCATACAAAGCAATCGTAAGCGCTGAAGCTGAAGAAA 2575795
QY 671 CCACCGTTGATAAGGCTGTCGATGTTGTTAGAGAGGATATGCTCAATCTGAGACTGACG 730
DB 2575794 -----TTTGAGATCCAATCGAACTTGTCAAAAGCTTCTTACGCTAATGACAAAATGATG 2575740
QY 731 AATTTCTGAACCAATTTGTTTTTCGGAGATGGGC-----GAGTAAAGATG 778
DB 2575739 AATTCGTTGTTCTCTGCTATCACTAAAGATGGCAACCTGTTGCAACAGTTAAAGACA 2575680
QY 779 AGGATACTCTTATTTTCTCAATATTCGTTGCTGATCGTATGCTGCTCAAAATTTGTTGATGTT 838
DB 2575679 ACGATGCACTGATTTTCTCAATTTCCGCTCGCTGCGTGCATTTCACTTTCTAATGCTAT 2575620
QY 839 TGGGCTCGAAGCTTATAAAGATCTTAATAGTTCGCTTCCCTCACCTCAAAAATATTCAGA 898
DB 2575619 TCACTGATAAAGATGGGATCATTTCCGACGCTGGGCTAACCCACCTAAAAACATCAAT 2575560
QY 899 TTAGTGGGATGACCAATACAAATAAAGAGCTTTCCATTTCCATCGTTATTTCCACCTGTGA 958
DB 2575559 TCGTTACATGACTCTTTACAAACCAAGCATTGATGCTGAAGTTGCTTTGAGCCANTTG 2575500
QY 959 CTCATACTAATGCTTGTGCTGAATGCTTCTCAAGGAGTTACTCAATTTCACTGCTG 1018
DB 2575499 AAATGAAAAATGTTATCGGCGAAGTACTTTCTTAATGAAGGTTTGTGCAACTGCTATCG 2575440
QY 1019 CGGAAACTGAGAGATGATCTCATGTTTACCTTCTTCTTAACTGGTGGTGCAGAGTTCAAT 1078
DB 2575439 CTGAAACAGAAAAATATCCACACGTAACATTTCTTATGAATGGCGACGAAATGAAGAAT 2575380
QY 1079 TCCAAGATGAAGCGTTGTTATGTTTCCGTACCAACAAAGAGTTGCTACATATGATTAA 1138
DB 2575379 TCCCTGGTGAACAAACCGGATTTCTAATCAATTCGCCAA---AAGTAGAACATACGATTAA 2575323
QY 1139 AACCAAGAAATGATGCTGCTGGAGTTGCCGAAAAAATGGTCGAGCAAAATTTGAGTCAGGCA 1198
DB 2575322 AACCTGAAATGATGCTATGATGAAGTAACAGATGCGCTGTTTGAAGACATTTAAAAACGACA 2575263
QY 1199 GGCATCCTTTGTTTATGTCGAATTTTGGCCCTCTGACATGTTGGACATCTGTTGAAT 1258
DB 2575262 AACAGATGCGATCATCTTAAACTTCGCTAATCCAGATATGTTGGACATTCAGGTATGC 2575203
QY 1259 TTGAACCTGCCCTCAAGCATGTCAAGCTACTGACGAGGCAATTTGAAAGATATTGGAAG 1318
DB 2575202 TTGACCAACTATTAAAGCAATCGAAGCATGATGAGTAAGTAATTTGGTGGTGTAGTAGTC 2575143
QY 1319 CATGCCAACTTATAATTTACGTTCTTATGGTTTACTTCCGATCATCGAAATGTGAGAAGA 1378
DB 2575142 TTATTTTAAAAAAGGTGGCTCAGCTATTATCTTGTGACCATGTTGACTCTGAAACAA 2575083
QY 1379 TGATTTGCTCCGATGGTAGTGAACATCTGACATACCTGCAATTTGGTCCCATTTACT 1437
DB 2575082 TGCTTACTCCAGAGGAAAAACCGCACACTGCGCACACTACCGTTCCAGTTCCAGTAAT 2575024

RESULT 4
ABR03041/c
ID ABR03041 standard; DNA; 2944528 BP.
XX
AC ABR03041;
XX
DT 05-FEB-2002 (first entry)
XX
DE Listeria monocytogenes EGD-e genome sequence.
XX
```

KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 XX vitamin B12; bacterial infection; disease; ds.
 OS Listeria monocytogenes.
 PN W0200177335-A2.
 XX 18-OCT-2001.
 XX 11-APR-2001; 2001WO-FR011118.
 XX 11-APR-2000; 2000FR-0004629.
 PR (INSP) INST PASTEUR.
 XX
 XX Bussieres C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
 PI Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;
 PI Daniels J, Goebel W, Kraft J, Kuhn N, Ng E, Vazquez-Boland JA;
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
 PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
 PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
 PI Rose M, Voss H;
 XX
 XX WPI; 2002-010914/01.
 DR
 XX
 XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
 PT and prevention of Listeria and related bacterial infections, and
 PT related polypeptides -
 XX
 XX Claim 1: SEQ ID No 1; 192pp; French.
 XX
 XX The present sequence is the genome sequence of Listeria monocytogenes
 CC EGD-e. This sequence and fragments of this sequence are useful for
 CC selecting probes and primers for detecting genes in L. monocytogenes and
 CC related organisms, and to study genetic polymorphisms and other genomes.
 CC Proteins (ABB47297-ABB50149) expressed from the present sequence are
 CC useful for raising specific antibodies, identification of L.
 CC monocytogenes and related organisms, and for biosynthesis and
 CC biodegradation, especially biosynthesis of vitamin B12. This sequence and
 CC proteins encoded by it are also useful for selecting compounds that
 CC regulate gene expression and cell replication and modulate L.
 CC monocytogenes-related diseases. In addition, this sequence and proteins
 CC encoded by it are useful in pharmaceutical and vaccines compositions for
 CC the treatment or prevention of infections by L. monocytogenes and related
 CC organisms.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;
 SQ
 Query Match 18.28; Score 312.6; DB 24; Length 2944528;
 Best Local Similarity 53.2; Pred. No. 4.8e-70;
 Matches 766; Conservative 0; Mismatches 649; Indels 24; Gaps 4;
 14 AGTTTGGAGTGGCAAAATATCAAAATGTTCAACAAAAGTCTGCTGTAGTTATTGATG 73
 2529205 ATTATTGAGGCACTAAATATGATGATGATGATGATGATGATGATGATGATG 2529146
 74 GATGGGCGCTTCCGATGACACACACCGGAATGCAATGCTAAAGCTAAACGCTATTA 133
 2529145 GTTTTGTAAACGTCGACAGACAGTAGTAAATGCTAGCTCAAGCAACAAACCAAA 2529086
 134 TGGACAACTTTGTTCTGGAATTTGGCA---AAATTTGGAGACACAGCTTTCATGTTG 190
 2529085 TCGACCTTTATTTGGCTGATTTTCTCCACGGGGAACCTTAAAGCTGCTGGCTTGTATG 2529026
 191 GATTGCCAAGAGGCTTAATGGGAATTTGGAATTTGGACATTTGAAATATAGGAGCTGAA 250
 2529025 GCCTTCAGAGGTCGAATGGGTAACTCTGAGTTGGCCATACAAACATCGGAGCTGGAC 2528966
 251 GAGTTATTTATCAAGATATTGTTGCAATTAATTTGGCTGTTCAACGAACGAGTTTGTGA 310

Db 2528965 GTATGCTACCAAGAGCTTAACCTGATTTGACAAAGCTATTGAAAGAGGCAATTCCAAG 2528906
 Qy 311 CAATCTCTAGATTTGTCATGAGCTGAGCGTGCAAGAGGAGGAGTGCTGATGCTATT 370
 Db 2528905 AGAACAAGCCCTAAACAATGCTTTTCACTATACAAAGAAACAACTCGACTTACATC 2528846
 Qy 371 TATTAGGACTGGTTAGCGATGGTGTCTCATATTTGATGATCATCTTTTTCGCTTGA 430
 Db 2528845 TTTTCGGTTTACTATATCAGACGGCGCTGATAGTACATTAATACCTCTGTCGACTTT 2528786
 Qy 431 TACGTGCATTTAAACAATTAACAAGTGCCTAAAGGTTTTCATTCACCTTTTTTCTGCTGATG 490
 Db 2528785 TAGAACTGCGAAAGATAAAGGCGTGAACGCTTATATCCATGATCTCTCTGATGAC 2528726
 Qy 491 GAGATACCTTCCCAACAAGTGGAGCTGGTTATCTTGAACAACCTTCTCAATTTATTCCTT 550
 Db 2528725 GTGACGCTGGGACCACTCTCTACATAGTAATTTAGAAACACACTCAAAAAGCTATTAGTG 2528666
 Qy 551 CGGAAAAGTACGGAGAATTTGGCTACTATTTACTGGACGTTATTTATGCAATGGATAGGACA 610
 Db 2528665 ACTTAACTATGGCGGATTTGCTACTGTTCTGACGCTTCTAGCGATGGATCGTATA 2528606
 Qy 611 AAAGATGGAGCGCTTAAAGATGGCTTTATGAGCAATTTGTTGGAGGTTATTTGGACAAAAG 670
 Db 2528605 AACGCTGGGAACGCTTTGAAAAGCATACAAAGCAATCGTAAGCGCTGAAGGTGAAAAA- 2528547
 Qy 671 CCACCGTTGATAGGCTGTCGATGTTGTTAGAGAGCGATGATGCTCAATCTGAGACTGACG 730
 Db 2528546 -----TTTGAAGATCCCAATGCTGCAAGCTTCTTATGCTAATGACAAAATATG 2528492
 Qy 731 AATTTCGAAACCAATTTGTTTTCGGACGATGGC-----GAGTAAAAAGATG 778
 Db 2528491 AATTCGTTGTTCTGCTATCATTAATAAGATGCAACCTGTTGCAACACTTAAAGACA 2528432
 Qy 779 ACAGTACTCTATTTTCTTCAATATGCTGATGCTGATGCTGATGCTGCAATTTGTTGAATGTT 838
 Db 2528431 ACAGTACAGTATTTTCTTCAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2528372
 Qy 839 TGGTCTCGAACGCTTAAAGATCTTAATAGTTCGTTTCTCCACCTCAAAAATATTTCAGA 898
 Db 2528371 TCACTGATAAAGAAATGGGATCATTTTCGACCGTGGGCGAGATCACCCCTAAAAACAAT 2528312
 Qy 899 TTAGTGGGATGACCAATACAAATAAGAGTTTCCATTTCCATGCTTTATTCACCACTGTGA 958
 Db 2528311 TTGTTACATGACTCTTTTCAACCCCAAGTATTGATGCTGAAGTTGCTTTTGAGCCAA 2528252
 Qy 959 CTCACTATATGCTGCTGATGATGGCTGCTTCTCAAGGAGTACTCAATTTTCACTGCTG 1018
 Db 2528251 AAATGAAAAATGTTATCGGCAAGTACTTTTCTAATGAAGGCTTATCGCACTGCTATCG 2528192
 Qy 1019 CGGAACTGAGAGTATCTCTCATGTTACCTTTCTTTTAAATGTTGCTGAGAACTTCAAT 1078
 Db 2528191 CTGAAACAGAAAAATATCCACACGTAACATTTCTTTTAAATGTTGAGGAGAAATGAAGA 2528132
 Qy 1079 TCCAAGATGAAGCGCTTGTATGCTTCCGTCACCAAAAGAGTTGCTACATATGATTTAA 1138
 Db 2528131 TCCCTGTGAAAAACCGGATTTCTAATCAATTCGCCAA---AAGTAAAAACATACGATTGC 2528075
 Qy 1139 AACCAAAATGAATGCTGCTGGAGTTGCCGAAAAAATGTTGCGAGCAAAATTCAGTCAAGCA 1198
 Db 2528074 AACCTGAATGAGCGCATATGAAGTAACAGATGCACTGTTGTAAGACATTTAAAAACGACA 2528015
 Qy 1199 GGCATCTTTGGTTATGTCGAATTTTGGCTCTCTGACATGTTGGGACATGCTGTAAT 1258
 Db 2528014 AACACATGCGATCATCTTAACTTCGCAACACGAGATATGTTGGGACCTCAGATGTC 2527955
 Qy 1259 TTGAACCTGCGTCAAAAGCATGCAAGCTACTGACGAGGCAATTTGGAAGAAATTTTGAAG 1318
 Db 2527954 TTGAGCCAAACGATTAAGCAATTCAGCACTAGATGAATCTTTGGTCTGTAGTACACC 2527895
 Qy 1319 CATGCCAACTTAAATTAATGCTTCTTATGTTTACTTCCGATCATGGAATGCTGAGAAGA 1378

Db	2527894	TTATTTTAAAAAGGTGGTTCACGATTATCTTTTGCTGACCATTGGTAACACTCTGAACA	2527835
Qy	1379	TGATGTGCTCCCGTAGTGAACATACACTGCACATACCTGCAATTTGGTCCCATTTACT	1437
Db	2527834	TGTCTACTCCAGAGAAACCGCACACTGCCACACTACCGTTCAGTTCACGATNTT	2527776
RESULT 5			
AAT58840/C	ID	AAT58840 standard; DNA; 580073 BP.	
XX	AAT58840;		
AC	27-MAR-1997	(first entry)	
DT	XX	Mycoplasma genitalium genome.	
DE	XX		
KW	M. genitalium; DNAB; DNA gyrase; origin of replication;		
KW	megabase shotgun sequencing method; open reading frame; ORF; ss.		
OS	Mycoplasma genitalium.		
FH	Key	Location/Qualifiers	
FT	CDS	8552..9184	
FT	/tag=	a	
FT	/label=	MG006	
FT	/note=	"Previously identified as MORF-20076, the encoded protein shows 27.59 percentage identity to thymidylate kinase (CDC8) from Saccharomyces cerevisiae"	
FT	11252..12040		
FT	/tag=	b	
FT	/label=	MG009	
FT	/note=	"Previously identified as MORF-20078, the encoded protein shows 35.43 percentage identity to the Bacillus subtilis hypothetical protein covered in accession number GB:D26185_102"	
FT	12069..12725		
FT	/tag=	c	
FT	/label=	MG010	
FT	/note=	"Previously identified as MORF-20079, the encoded protein shows 25.73 percentage identity to DNA primase (dnaE) from Clostridium acetobutylicum"	
FT	complement (13570..14247)		
FT	/tag=	d	
FT	/label=	MG012	
FT	/note=	"Previously identified as MORF-20080, the encoded protein shows 31.50 percentage identity to the ribosomal protein S6 modification protein (rimK) from Escherichia coli."	
FT	complement (14396..15217)		
FT	/tag=	e	
FT	/label=	MG013	
FT	/note=	"Previously identified as MORF-19823, MORF-20080 and MORF-20081, the encoded protein shows 33.04 percentage identity to 5,10-methylene-tetrahydrofolate dehydrogenase (fold) from E. coli"	
FT	17474..19243		
FT	/tag=	f	
FT	/label=	MG015	
FT	/note=	"Previously identified as MORF-20084, the encoded protein shows 32.23 percentage identity to transport ATP-binding protein (msbA) from E. coli"	
FT	26478..27344		
FT	/tag=	g	
FT	/label=	MG023	
FT	/note=	"Previously identified as MORF-20092, the encoded protein shows 45.96 percentage identity to fructose-bisphosphate aldolase"	

Db 536434 ATGCACACTAGAAAAAGCTTAA-----TAACCTTTGATTCTTACTGTTTAAATTTTGCTA 536381
 QY 1229 CTCTCAGCATGGTTGGACATACACTGTAATTTGCAACCTGCCGTCAAGCATGTCAGACTA 1288
 Db 536380 ATCCTGATAGTGGTACATCTGCTGTAACCTATCAAGCTTGCATTAAGGCTCTTGAAGCAC 536321
 QY 1289 CTGACGAGGCAATTTGGAAGATATTTGAAGCATGCCAACTTATAATTAAGCTTCTTTATGG 1348
 Db 536320 TCGATGTTCAAATTTAAACGAATAGTTGATTTTGTAAAGCTATCAATAAATCACTATCTTT 536261
 QY 1349 TTAATTCCTGATGTAAGTGTGAGCAAGATGTTCTCCGATGTTAGTGAACATACG 1408
 Db 536260 TAACTGCAGATCATGCGAATGCGAAGTGTGATGATGATTAATAACAATCCAGTTACTA 536201
 QY 1409 CACATACCTGCAATTTGGTCCCATTTACTTGTCTTCCCAACATTTGTTTAAATCGA 1468
 Db 536200 AACACACTATTATCTCTGATACCATTTGTATGACTGACAAAAATGTTAACTTTAATCAAA 536141
 QY 1469 CT 1470
 Db 536140 CT 536139

RESULT 6

ABN93144
 ID ABN93144 standard; DNA; 1554 BP.
 XX AC
 XX ABN93144;

DT 24-JUL-2002 (first entry)

DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:2607.

DE Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 KW antibacterial; gene therapy; gene; ds.

XX Staphylococcus epidermidis.

OS US6380370-B1.

PN 30-APR-2002.

XX 13-AUG-1998; 98US-0134001.

PR 14-AUG-1997; 97US-055779P.

PR 08-NOV-1997; 97US-064964P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Doucette-Stamm LA, Bush D;

XX WPI; 2002-381255/41.

DR P-PSDB; ABP40599.

PT Novel isolated nucleic acid encoding a Staphylococcus epidermidis
 PT polypeptide, useful for diagnosing and treating bacterial infections -
 XX Disclosure; SEQ ID 2607; 267pp; English.

XX ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
 CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
 CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
 CC antibacterial activity and can be used in gene therapy. The sequences
 CC can also be used in the diagnosis and treatment of bacterial infections,
 CC particularly S. epidermidis infections. The sequences can be used to
 CC screen for compounds able to interfere with the S. epidermidis life
 CC cycle or inhibit S. epidermidis infection.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site.

XX Sequence 1554 BP; 538 A; 240 C; 321 G; 455 T; 0 other;

Query Match 15.8%; Score 271.2; DB 24; Length 1554;
 Best Local Similarity 51.4%; Pred. No. 1.6e-60;
 Matches 759; Conservative 0; Mismatches 703; Indels 16; Gaps 5;
 QY 12 CAAGTTTGATGAGGACAAATATCAAAATGTTCAACAAAAAGTCTGTCTT-GTAGTTATTG 70
 Db 11 CAATTGTTAGAAGGTGCAAAATAAATATGCGAAAAACAACCACTGCTTAATCATCTTAG 70
 QY 71 ATGGATGGGCGCTTCCCGATGACACACGCGGAATCAATTTGCTAAGCTAAACGCCCTA 130
 Db 71 ATGGTTTCCCAATTCGTGAAGTGAACATGCGAATGAGCAATAGCAAGCAATCAAAACCTA 130
 QY 131 TTATGACAAACCTTTTCTTGGAAATTTGGCAAA--AATTGGAAGCACACGGTCTTTCATG 187
 Db 131 ATTTTGATCGATATTATGAAAAATATCCTACAACACAAATAGAAGCTAGTGGCTTAGATG 190
 QY 188 TTGGATTGCCAAGAGCTTAATGGGAAATTCGAAGTTGGACATTTGAATATAGGAGCTG 247
 Db 191 TAGGTCTTCTGAAGGTCAAAATGGGTAACTGTAAGTAGGACATATGAATATTTGGTGCAG 250
 QY 248 GAAGAGTTATTATCAAGATATTGTTCTGAATTAATTTGGCTGTTCAACGAAAGGACTTTG 307
 Db 251 GACGATCGTATATCAAGTTTAACCTGATTAATAAATCGATTGAAGCGGGAATTTCT 310
 QY 308 TTACAAATCCTCAGATTTGTCATCAGCTGAGCGTCAAGAGGAGGAGTGTGTCGATTCG 367
 Db 311 TTGATAACACACTGTATTAAATAACACACTGTTAAACATGTTAAAGACAATGCTCTCGCCTTC 370
 QY 368 ATTTATTAGGACCTGTTAGCGATGTTGGTGTCCACTCTCATATTGATCATCTTTTTCGCTG 427
 Db 371 ATGTATTCGGATTGCTTCTGATGGTGTGTACACAGTCATTTATAAGCATCTATTTCCTA 430
 QY 428 TCATACGTGCATTTAAACAATTTACAAGTCCCAAGGTTTTCATTCACCTTTTTCGCTGATG 487
 Db 431 TTTTGAATTAAGTCTAAAAAGCAAGGATAGATATAAGTATATGTCCACGCAATTTTAGATG 490
 QY 488 GTCGAGATACCTTCGCCAACAAAGTGGAGCTGGTTATCTTGAACAACCTTCCTCAATTTATTG 547
 Db 491 GTCGTGATGTTGATCAAAATCTGCTTTGAAATATATAGAGGAAACTGAAGATAAATTTA 550
 QY 548 CTTTCGAAAGTACGGAGAAATTTGGCTACTATTACTGGAGCTTATTATGCAATGATAGG 607
 Db 551 AAGAAATAGGTGATAGGCCAATTTCCGTTTCAGGACGTTATTATGCTATGGACCGTG 610
 QY 608 ACAAAAGATGGAGGCGTATTAAAGATGCTTTATGAGGCAATTTGTTGAGGCTATTGGACAAA 667
 Db 611 ACAAGCGTTGGGATCGTGAGGAACGTCCTATAATGCTATTTCGTAACCTTTGAAGGCTCTA 670
 QY 668 AAGCCACCGTTGATTAAGGCTGTGATGTTGTTAGAGAGCGATATGCTCAATCTGAGACTG 727
 Db 671 CATTTACTTCAGCTAAAGCAGCGCTTGAAGCTTAATTAATAAATGATGTGACTGATGAAT 730
 QY 728 ACGAATTTCTGAAACCAATTTGTTTTCGGACGATGGGCGAGTAAAGATACGATACATC 787
 Db 731 TCGTCGACCGTTTATAGTTGAAGGCCAAACAGTGGT---GTGAACGATGAGACGCCAG 787
 QY 788 TTATTTTCTTCAATTTATCGTGTATCGCTATCGCTCAAAATTTGTGAATTTGTTGGGCTCG 847
 Db 788 TAATCTTTTATAATTTCCGTCAGATAGACGAGCTCAACTTTTCAGAAATCTTTTACTAATA 847
 QY 848 AACGTTATAAGATCTTAAATAGTTCGTTCCCTCACCTTAAATAATTTAGATTTAGTGGGA 907
 Db 848 AAGCGTTTGTATGATTTTAA-----AGTTGAACAAGTGGACAACCTATTCTACGCTACAT 901
 QY 908 TGACCCAATACAAATAAAGATTTTCCATTTTCCATCTGTTATCCACCTGTGACTCATACTA 967
 Db 902 TCACCAATATATACATGTAGATGCTGAATTTGTTATTTCAAAAAGTTGACTTAATAA 961
 QY 968 ATGTCCTTCTGTAAGTGGTTCCTTCAAGGAGTTTACTCAATTTTCACTGTGCGGAACATG 1027
 Db 962 ATACAATCGGTGAAGTTGCTCAAGATAAATGAGTGTGAAACAATTTACGCTACGCTGAAACTG 1021
 QY 1028 AGAAGTATCCCTCATCTTACCTTCTCTTTAATGGTGGTGGAGAAAGTTCAATTTCCAAGATG 1087

Db 1567 TAATCTTTTATATTTCCCTCCAGATAGACAGCTCAACTTTTCAGAAATCTTTACTAATA 1508
QY 848 ACGTTTATAAGATCTTAATAGTTTCCTCCACCTTAAATAATTTAGATTAGTGGA 907
Db 1507 AAGCGTTTATGGATTAA-----AGTTGAACAAGTGGACACATTTTCTAGGCTACAT 1454
QY 908 TGACCAATACATAAAGAGTTTCCATTTCCATGTTATTTCCACCTGTGACTCATACATA 967
Db 1453 TCACGAAATATATGACATGATAGATGCTGAAATTTGTTGAAAAGTTGACTTTAAATA 1394
QY 968 ATGTGCTTCTGAATGGCTTCTTCTCAGGAGTTTACTCAATTTCTACTGTGGGAACTG 1027
Db 1393 ATACAATCGGTGAAGTTGCTCAAGATTAATGGTTTGAACAATTTACGTATCGTGAACGTG 1334
QY 1028 AGAAGTATCTCATGTACTTCTTCTTTAAATGGTGTGCGAGAGTTCAATTTCCAAAGATG 1087
Db 1333 AAAAGTATCCACATGTAACATCTTATGAGTGGGAGCAATGAGAGTTTGAAGGAG 1274
QY 1088 AAGAGCGTTGTATGGTTCCGTCAACCAAAAGAGTTGCTACATATGATTTAAACCCAGAA 1147
Db 1273 AACGTGCTAGACTCATCGATTCTCCAA---AAGTAGCGACTTATGATTTAAACCTGAGA 1217
QY 1148 TGAATGCTCTGAGTTGCGGAAATAATGGTCGACCAAAATTTGAGTCAGCGAGCATCCTT 1207
Db 1216 TGAGTGCATATGAAGTTAAAGATGCAATTTATGAGAGGTAGACAAAGGTGACTTAGATT 1157
QY 1208 TGGTTATGCAATTTTGGCCCTCTGACATGTTGGACATATCTGTAATTTGAACCTG 1267
Db 1156 TAATCTACTGAATTTGCTTAACCCAGATATGTTGGACATATGTTGATGCTTGAACCAA 1097
QY 1268 CCGTCAACCATGTCAGTACTGACGAGGCAATTTGGAAGATATTTGAAGCATGCCAA 1327
Db 1096 CAATTAAGCAATGAAGAGTATGATGAGTCTTGGTGAAGTCTGTGACAAAATTTATTG 1037
QY 1328 CTTATAATTTACGTCTTATGTTTACTTCCGATCATGGAAATGCTGAGAGATGATGCTC 1387
Db 1036 ATATGGTGTGTCATGCCATCATCTGTCAGACACCGGTAACTCAGATTAATTAACGTG 977
QY 1388 CCGATGGTATGAACATATGCAATGCTGCAATTTGGTCCCATTTACTTCTCTTCCA 1447
Db 976 ATGACGACCAACCTATGACGACACACACAACTTAATCTGTTCCAGTATTTGAACCTAAG 917
QY 1448 AAACATTTGTTTTTAATGCACTCCACCTACTGGAGAT 1485
Db 916 AAGGTGTTACATTAAGAGAAACTGGACGTTTAGCGCAT 879

RESULT 8

AAH53413
XX ID AAH53413 standard; DNA; 1518 BP.
XX AC AAH53413;
XX DT 03-SEP-2001 (first entry)
XX DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:2219.
XX DE Staphylococcus epidermidis SR1 strain; infection; diagnosis;
XX KW vaccination; endocarditis; ds.
XX OS Staphylococcus epidermidis.
XX PN WO200134809-A2.
XX PD 17-MAY-2001.
XX PF 09-NOV-2000; 2000WO-US30782.
XX PR 09-NOV-1999; 99US-0164258.
XX PA (GLAXO) GLAXO GROUP LTD.
XX

PI Kimmerly WJ;
XX WPI; 2001-316495/33.
DR P-PSDB; AAG82563.
XX
XX
PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
PS
PS Claim 8; Page 594-595; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
XX Sequence 1518 BP; 524 A; 236 C; 315 G; 443 T; 0 other;

Query Match 15.5%; Score 266.2; DB 22; Length 1518;
Best Local Similarity 51.0%; Pred. No. 3.2e-59;
Matches 737; Conservative 0; Mismatches 593; Indels 15; Gaps 4;
QY 44 AACAAAAAGTCTGCTTGTAGTTATTTGATGGATGGGGCCCTTTCCGATGAACACGCGGA 103
Db 8 AACAAACCACTCCCTTAATCATCTTAGATGGTTTCCGAAATCGTGAAGTGAACATGGCA 67
QY 104 ATGCAATTCGTAAGCTAAACGACGCTTATTTATGACAAACCTTTGTTCTGCAATTTGCCAA 163
Db 68 ATGCAATTAAGCAACACATAAACCTTATTTGATCGATATTTATGAAATATCTCTACAA 127
QY 164 ---AATTTGAAGACACACGCTCTTCATGTTGGATTTCCAGAGGCTTAATTTGGAAATTTCTG 220
Db 128 CACAATAGAACCTAGTGGCTTAGATGATGTTCTTCTGAAAGTCAATTTGGTAACTCTG 187
QY 221 AAGTTGGACATTTGAATATAGAGCTGGAGAGTTATTTATCAAGATATTTGTTTCAATTA 280
Db 188 AAGTAGGACATATGAATATTTGTCGAGGACGATCTATATCAAAAGTTTAACTCGTATTA 247
QY 281 ATTTGGCTGTTCAACGAAACGAGTTTGTATCAAAATCTTCAGATTTGTCATCAGCTGAGC 340
Db 248 ATAAATCGATTGAAGACGAGGAATTTCTTTGATAACACTGATTAATAACGCTGTTTAAAC 307
QY 341 GTGCAAGAAGGGGAGTGGTTCGATTTTATAGGACTGTTAGCGATGTTGGTGTCC 400
Db 308 ATGTTTAAACAAATGCTCTCGCTTCATGATTTCCGATTTCTTCTGATGTTGGTGTAC 367
QY 401 ACTCTCATATGATCATCTTTTGGTGGTGTGATACGTGCAATTTAAACAATTTACAAAGTCCAA 460
Db 368 ACAGTCATTAAGCATCTATTTTGTCTATTTTGAATTTAGCTAAAAAGCAAGGAATAGATA 427
QY 461 AGGTTTTCATTCACCTTTTTCGCTGATGTCGAGATCTTCGCAACAGTGGAGCTGGTT 520
Db 428 AAGTATATGTCACCGCAATTTTATGATGGTGGTATTTGATCAAAAATTTGCTTGAAT 487
QY 521 ATCTTGAACAACTTCTTCAATTTATTTGCTTTCGAAAAAGTACGAGAAATTTGGCTACTATTA 580
Db 488 ATATAGAGAACTCAACATAAATTTAAAGATTTAGGTCTAGGCCAATTCGCTTCTGTTT 547

Matches 709; Conservative 1; Mismatches 701; Indels 22; Gaps 6;

QY	12	CAAGTTTGAGATGGACAAATATCAAAATGCTTCAACAAAAGGCTGCTTT-GTAGATTATG	70
Db	4390	CAATTGTTAGAGGTGCAAAATTAATCATGGCTAAGAAACCAACTCGGTTAATTAATTAG	4449
QY	71	ATGGATGGGCGCTTCCCGATGAACACACGGGAATGCAATTGCTAAAGCTAAACGCCCTA	130
Db	4450	ATGGTTTTCGGAACCGCAACGCAACATGTAATGCGGTAAATTAGCAACAGCCTA	4509
QY	131	TTATGACAAACTTTGTTCTTGGAAAT---TGGCAAAATTTGAAGACACAGCTCTTCATG	187
Db	4510	ATTTTTCATCGTTTATTAACAAATATCCAACGACTCAATTCGAACGGATGGCTTAGATG	4569
QY	188	TTGGATTGGCAAGCGCTTAATGGGAATTCGAAAGTTGGACATTTGAATATAGAGGTG	247
Db	4570	TTGGACTACCTGAAGCAAAATGGGTAACTCAGAAGTTGGTCATATGATATCGGTGAG	4629
QY	248	GAAGAGTTTATTTATCAAGATATGTTTCAATTTTGGCTGTTTCAACGAAACGAGTTTG	307
Db	4630	GACGTATCGTTTATCAAAAGTTTAACTCGAATCAATAAATCAATTCGAAGACGGTATTCT	4689
QY	308	TTACAATCCTCAGATTTGTCATCAGCTGACGTGCGTGCRAAGAGGGAGTGCATATGC	367
Db	4690	TTGAAATATGATGTTTAAATATATGCAATTCACACGTGAATTCACATGATTCAGCGTTAC	4749
QY	368	ATTTATTAGGACTGCTAGCGATGGTGGTGCTCCACTCATATTTGATCATCTTTTGGCT	427
Db	4750	ACATCTTTGGTTTATGCTGACGGTGGTGACAGCTATTCACAAACATTTATTGGCTT	4809
QY	428	TGATACGTGCAATTAACAATTTACAAGTGCCAAAGGTTTTCATTCACTTTTTGGTGATG	487
Db	4810	TGTTAGAATCTGCTTAAACAAACAGGTGTGAACAAAGTTTACGTACACGCTTTTATGATG	4869
QY	488	GTCGAGATCTTCGCCAACAGTGGAGCTGGTTATCTTGAACAACTTCTCAATTTATG	547
Db	4870	GCCGTGACGTAGATCAAAATCCGCTTTGAAATACATCGAAGAGACTGAAGCTAAATCA	4929
QY	548	CTTCGGAAGTACGGAATTTGGCTACTATTACTGGAGCTTATTATGCAATGGATAGG	607
Db	4930	ATGAATTAGGCATTGGTCATTTGGCATCTGCTGCTGCTGCTTATTATGCAATGGATCGTG	4989
QY	608	ACAAAGATGGGAGCGTATTAGATGGCTTATGAGGCAATTTGGAGGTATTTGACAAA	667
Db	4990	ACAAACGTTGGACGCTGAAGAAAAGCTTACAATGCTATT-----CGTAATTTTGATG	5043
QY	668	AGCCACCGTTCATAGGCTGTCGATGTTGTTAGAGAGCGATATGCTCAATCTGAGACTG	727
Db	5044	CCCCAATATGCAACTGCCAAAGAGGTGTAGAAGCAAGCTTAAATGAGGCTTTAACTG	5103
QY	728	ACGAATTTCTGAACCA---ATTGTTTTTTCGGACGATGGGCGAGTAAAGATGACGATA	784
Db	5104	ACGAATTCGTAGTACCATTCATCGTTGAGAATCAAAATGACGGTGTAAATGATGAGNNN	5163
QY	785	CTCTTATTTCTCAATTATCGTGTGATCGTATCGGTCAAAATTTGTGAATGTTTGGGTC	844
Db	5164	NN	5223
QY	845	TCGACGTTATAAAGATCTTAATAGTTCCGTTCTCACCTTAAATAATATTCAGATTAGTG	904
Db	5234	ACAGAGCATTCGAAGCGCTTTAAA-----GTTGAACAAAGTTAAAGACTTATCTATGAA	5277
QY	905	GGATGACCAATACATAAAGAGTTTCCATTTCCATCGTTATTCACCTGTGACTCATATA	964
Db	5278	CATTCACATAGTATATGACATATCGATGCGGCTATCGTCTTCGAAAAGTTGATTTAA	5337
QY	965	CTAATGCTGCTGCAATGGCTTCTTCTCAAGGAGTTACTCAATTTTCACTGTGCGGAAA	1024
Db	5338	ATAATACAATTTGGTGAATTTGCACAAAATAACAATTTAACTCAATTTACGTATTGCAGAA	5397
QY	1025	CTGAGAAGTATCCTCATGTTACCTTCTTCTTATGTTGGTGGCGAGAAGTTCAATTCGAAG	1084
Db	5398	CTGAAAATAACCCCTCACGTTACTTACTTATGAGTGGTGGAGCTAACGAGGAATTTAAAG	5457

1085 ATGAAGAGCGTTGTATGTTGCTTCGTCACCAAAAAGAGTTGCTACATATGATTTAAACACAG 1144
1144
5458 GTGAACGCGCTGTTTAATTTGATTCACCTA---AAGTTGCAACGTATGACTTGAACACAG 5514
5514
1145 AAATGAATGCTGCTGGAGTTGCCGAAAAAATGTCGAGCAAAATGAGTCAGCGAGGCATC 1204
1204
5515 AAATGAGTCTTATCAAGTTAAAGATGCAATTTATGAAGAGTTAAATAAAGCTGACTTGG 5574
5574
1205 CTTTGGTTATGTCGAATTTTGGCCCTCCTGACATGTTGGACATACTGGTAAATTTGAAC 1264
1264
5575 ACTTAATATTATTAAACCTTGTCTAACCTGATATGTTGGACATAGTGTATGCTGTGAGC 5634
5634
1265 CTGCGCTCAAAAGCATGCTCAAGCTACTGACGAGGCAATTTGGAAGAGATATTTGAAGCATGCC 1324
1324
5635 CGACAATCAAGCAATCGAAGCGTTGATGAATTTTAGGAGAAGTGGTTGATAAGATT 5694
5694
1325 AAATTAATTAATACCTTCTTATGTTTACTTCCGATCATGGAATGCTGAGAAGATGATG 1384
1384
5695 TAGACATGGAGCGTTTATGCAATTTATCTGCTGACCATGGTAACTCTGATCAAGTATTGA 5754
5754
1385 CTCGCGATGGTAGTGAACATACTGCACATACCTGCAATTTGGTCCCATTTACT 1437
1437
5755 CGATGATGATCAACCAATGACTACGAWACAACCAACCCAGTACCATGATT 5807
5807

RESULT 10
AAF28544
ID AAF28544 standard; DNA; 65792 BP.
XX
AC AAF28544;
XX
DT 04-APR-2001 (first entry)
XX
DE Genomic fragment #31.
XX
KW Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW bronchopulmonary; endocarditis; meningitis; ss.
XX
OS Moraxella catarrhalis.
XX
PN WO200078968-A2.
XX
PD 28-DEC-2000.
XX
PF 16-JUN-2000; 2000WO-US16649.
XX
PR 18-JUN-1999; 99US-0140121.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Lagace RE, Patterson C, Berg KL;
DR WPI; 2001-041427/05.
XX
PT Genomic library for identifying diagnostic and therapeutic
PT compositions, and for identifying virulence factors, regulatory
PT elements and drug targets, comprises Moraxella catarrhalis nucleic
PT acids
XX
PS Claim 1; Page 278-293; 545pp; English.
XX
CC The present invention relates to a Moraxella catarrhalis genomic library
CC comprising of a combination of 41 nucleic acid molecules (see
CC AAF28544-AAF28554). The library has a number of uses described in the
CC specification e.g. is useful for identifying diagnostic and therapeutic
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC aerobic, gram-negative diplococcus, normally found among the bacterial
CC flora of human upper airways. M. catarrhalis is known to cause acute,
CC localised infections such as otitis media, sinusitis and bronchopulmonary
CC infection and life-threatening, systemic diseases including endocarditis
CC and meningitis.
XX

CC infections, including acute, chronic, and atrophic gastritis, and peptic
CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC for the production of antibodies. The products can also be used for
CC detection and diagnosis.

XX Sequence 1620 BP; 518 A; 324 C; 348 G; 430 T; 0 other;

Query Match 12.9%; Score 222.2; DB 19; Length 1620;
Best Local Similarity 53.9%; Pred. No. 1e-47;
Matches 655; Conservative 0; Mismatches 518; Indels 42; Gaps 8;

QY 167 TGAAGCAGACGGCTCTTCATGTTGGATTCGCGAAGGCTTAATGGAATTCGAAAGTTG 226
DB 196 TTGATACGATGCTTTCAGCGTGGCTTACCTAAGGGGCAAAATGGAATTCGAAAGTTG 255
QY 227 GACATTTGAATAGAGCTGGAAGAGTTATATCAAGATATGTCGATTAATTTGG 286
DB 256 GGCATATGTTATGCGGCTGGTGGTGGTCTATCAGGATTTAGTCAAAATTTCTTTAA 315
QY 287 CTGTTCAAGCAACAGGTTTGTACAAATCCTCAGATTTGTCATCAGCTGAGCGTCA 346
DB 316 GCCTTCAACAGATGAATTAACAAACACCCCGCTTTT-----AACACGATCCAA 368
QY 347 AGAAGGGAGTGTGCGATTCATTTATAGGACTGTTAGCGATGTTGGTTCACCTCTC 406
DB 369 AAAAGCCCTGTG-----TGCATCTTATGGTTTAAATGAGCGATGAGCGGTGCAATCAC 423
QY 407 ATATTGATCATCTTTTGGCTGTGATGCTGCAATTTAAACAAATACAAAGTTC 466
DB 424 ACATTGAGCATTTTATCGCTCTGCTGTTTAGAGTGTGAAAATCCCA---TAAAAAAGTCT 480
QY 467 TCATTCACTTTTGTGCTGATGTCGAGATATTCGCAACAAAGTGGTGTATCTTG 526
DB 481 GTCTGATTTAATCAGGATGGCGGATGTCGCTCTAAAGCGTTTAACTATTATA 540
QY 527 ACAACTCTTCAATTTATGCTTCGGAAGTACGGAATTTGGTACTATTACTGAGAC 586
DB 541 AACAAATGC-----AAAATATCTGCAATGAAGAGTTCAAATTCGCTACCAATGAGGCTC 594
QY 587 GTTATTATCAATGATAGGCAAAAGATGGAGGCTTAAAGATGGCTTTATGAGGCA 646
DB 595 GTTTTATGCCATGATAGGATAGCGCTTTGAAGAGTGTGAGCTTGGCTGTCA----- 649
QY 647 TTGTTGGAGTATTTGACAAAAGCCGTTGATAAGCTGTGCGATGTTGTAGAGAC 706
DB 650 ----TAGCTTAATGGGCTTAATCAACAGCTTTAAGCCCTAGCGATATATCCAAAGCC 705
QY 707 GATATCTCAATCTGAGACTGAGGATTTCTGAAACCAATTTGTTTTCGGAGGAGG 766
DB 706 AGTATGATAAAATATCACCGATGAATTTATCATGCCCGCTTGTTTTAAAAAATTTATGCG 765
QY 767 GAGTAAAGATGACGATACCTTATTTCTTCAATTTATGCTGATGCTATGCGTCAAA 826
DB 766 GCATGCAAGATGATGAGAGTTTATTTTATCAATTTAGGATGATAGGCTAGAGAA 825
QY 827 TTTGTGAATTTTGGCTCTCGAACGTTTAAAGATCTTAATAGTTTCGTTCTCACCCCTA 886
DB 826 TCGTGAGCGCTTTAGGCCAAAACAAATTCAGTGGCTTTAAGCGCAAGTT-----TTTA 879
QY 887 AAAATATTAGATTTAGTGGATGACCCATACAAATGAAGAGTTTCCATTTCCATCGTTAT 946
DB 880 AAAAATCCATATCGCTACCATGACGCTTATGATGAACACTTTTCCCTACCTCGTTTAT 939
QY 947 TCCACCTGTGACTCATACTATGTTGCTGCTGAATGGCTTCTCTCAAGGAGTTACTTC 1006
DB 940 TCCCAACAAAGCGTTCAAAACAGCTCGCTGAAGTGTCTCTCAACACACCTGACCC 999
QY 1007 AATTTACTGTGCGGAACCTGAGAAGTATCTTCATGTTTACCTTCTTTTAAATGTTGTC 1066
DB 1000 AAAGCCATATCTGTAACCTGAAATGAAATACGCGCATTAACCTTTTTCATCATGCGGAG 1059
QY 1067 GAGAAGTTCAATCCAGATGAGAGCGTTGTATGTTCCGTCACCAAAAGAGTGTCTA 1126

DB 1060 TGGAGACGCGCTTTTAAAAATGAAACCCGGTGCTTATCCAAAGCCCTA---AAGTTACCA 1116
QY 1127 CATATGATTTAAACCAGAAATCAATGCTGCTGGAGTTCCGAAAAAATGGTCGACAAA 1186
DB 1117 CTTATGACTTAAGCCTGAATGAGCGCTAAAGAGTAACCCCTGGGCTGTTAGACAAA 1176
QY 1187 TTGAGTCAGCAGGCGATCCTTTGGTTATGTGCAATTTTGGCTCTCTGACATGGTTGGAC 1246
DB 1177 TGAACACTAGCAGCGA---TTTGATCATTTGTAATTTTCTATGCGGATATGTTAGGCG 1233
QY 1247 ATACTGGTAAATTTGAACCTGCGCTCAAGCATGCTCAAGTACTGACGAGGCAATTTGAA 1306
DB 1234 ATACGGGGAATTTGAAGCGAGCGTCAAGCGGTGAAGCAGTGGATGATGTTTAGGG 1293
QY 1307 AGATATTTCAAGCATGCCAAACTTATAATAGTCTTCTTATGTTACTTCCGATCATGGAA 1366
DB 1294 AATCTCTTCACTGGCTAAATAATTTGATTAACCCATGCTTTTAAACGAGCATCATGGGA 1353
QY 1367 ATGCTGAGAAGATGA 1381
DB 1354 ATTGCGAGCGCATGA 1368

RESULT 12

ABS53606/C

ID ABS53606 standard; DNA; 11769 BP.

XX ABS53606;

DT 29-NOV-2002 (first entry)

XX Human Mut/NUDIX DNA sequence.

XX Human; ds; Mut/NUDIX; housecleaning enzyme; cancer;
XX nucleoside triphosphate pyrophosphorylase; cardiovascular disorder;
XX central nervous system disorder; congestive heart failure; ischaemia;
XX arrhythmia; hypertensive vascular disease; peripheral vascular disease;
XX myocardial infarction; angina; brain injury; mood disorder; gene therapy;
XX anxiety; myopathic disorder; neurodegenerative disease; chronic pain;
XX Alzheimer's disease; Parkinson's disease; multiple sclerosis;
XX epilepsy; acquired immunodeficiency syndrome-related pain; migraine;
XX dementia; leukaemia; lymphoma.

XX Homo sapiens.

XX WO200266626-A2.

XX 29-AUG-2002.

XX 14-FEB-2002; 2002WO-EP01541.

XX 16-FEB-2001; 2001US-268862P.

XX (FARB) BAYER AG.

XX Smith TJ;

XX WPI; 2002-674936/72.

XX New human Mut/NUDIX nucleoside triphosphate pyrophosphorylase
XX (Mut/NUDIX NTP) and its encoded protein, useful for identifying
XX modulators of Mut/NUDIX NTP activity, and in gene therapy for treating
XX cancer or heart disease.

XX Disclosure; Fig 3; 135pp; English.

XX The invention relates to an isolated polynucleotide, which encodes a
XX human Mut/NUDIX nucleoside triphosphate pyrophosphorylase, a
XX "housecleaning" enzyme. Also included are Mut/NUDIX expression
XX vectors, host cells, fusion proteins, antibodies and methods of screening
XX for agents that modulate (increase or decrease) the activity of
XX Mut/NUDIX protein or nucleic acid. The Mut/NUDIX NTP polynucleotide and
XX polypeptide are useful for identifying test compounds, that may act as

